[16] Hurwitz, J., A. Bresler, and R. Dinger, *Biochem. Biophys. Research Commun.*, **3**, 15 (1960).
[17] Stevens, A., *op. cit.*, **3**, 92 (1960).
[18] Furth, J. J., J. Hurwitz, and M. Goldmann, *op. cit.*, **4**, 362 (1961).
[19] Weiss, S. B., and T. Nakamoto, these PROCEEDINGS, **47**, 694 (1961).
[20] Ochoa, S., D. P. Burma, H. Kröger, and J. D. Weill, *op. cit.*, **47**, 670 (1961).

CORRELATION BETWEEN BASE COMPOSITION OF DEOXYRIBONUCLEIC ACID AND AMINO ACID COMPOSITION OF PROTEIN

BY NOBORU SUEOKA

DEPARTMENT OF MICROBIOLOGY, UNIVERSITY OF ILLINOIS

Communicated by A. D. Hershey, June 1, 1961

It has been known that the base composition of deoxyribonucleic acid (DNA) is quite variable from organism to organism. A good example of this variability is provided by bacterial species. Here, the mean guanine-cytosine content ranges from 26 to 74 per cent.[1, 2] Moreover, the heterogeneity of base composition of DNA molecules from each species is relatively small.[3–6] These observations would appear to provide a fruitful basis for comparative biological investigations in molecular terms.

In the current views of DNA-protein coding, it is assumed that the base sequence of DNA has a direct correspondence with amino acid sequence in protein. However, whether the implied correlations between DNA base composition and amino acid composition of protein exist or not depends upon the mechanism of coding, its universality, and the magnitudes of the participating fractions of DNA and protein in the coding relation. If the majority of proteins do have coding relations with DNA, and if, in particular, the code is universal and four-symboled, we may expect some correlations between DNA and protein in their composition.

From these points of view, a search for possible correlations between DNA base composition and amino acid composition of protein has been initiated. In the present paper, the results of the first stage of our work which is concerned with total protein (microsomal plus soluble) of bacterial cells will be presented. Amino acid compositions of bulk protein preparations isolated from 11 bacterial species whose DNA base composition varies from 35 to 72 per cent in guanine-cytosine content have been examined. The results indicate that correlations between component compositions of DNA and protein do indeed exist.

Materials and Methods.—Most bacterial species used in the present investigation have been obtained from the American Type Culture Collection. One liter of an enriched nutrient broth medium[7] was inoculated and incubated at 37°C for about 15 hr with constant shaking. The minimal medium used was *M* 26 basal minimal medium.[8] This medium was supplemented with glutamic acid (0.5 gm per liter) for *Bacillus subtilis* and *Serratia marcescens*. Cells were concentrated by centrifugation at 4,000 *g* for 15 min at 5°C, washed once with tris magnesium buffer (10^{-2} *M* Tris (hydroxymethyl) aminomethane plus 10^{-3} *M* Mg, pH 7.3), and suspended in the same buffer to make the final volume 20 ml. In order to disrupt the cells, the cell suspension was put through a French press twice under a pressure of 12,000 lb/inch^2, keeping the system as cold as possible. The resulting suspension was centrifuged at 6,000 *g* for 15 min at 5°C, and the

TABLE 1 AMINO ACID COM-

Strains§	*Bacillus cereus*		*Bacillus megatherium*	*Bacillus subtilis*			*Salmonella typhimurium*	*Escherichia coli*			
DNA GC* Content (%)	ATCC 6464 35		KM 38	NCTC 3610 42			50	B 50			
Medium†	EBM Prep 2	EBM Prep 2	EBM Prep 2	EBM Prep 2	EBM Prep 2	MM Prep 2	EBM Prep 2	EBM Prep 1	EBM Prep 2	EBM Prep 2	MM Prep 2
					Stable						
Lys	6.8	10.4	10.1	11.6	9.4	10.0	9.5	8.9	8.9	8.9	8.5
His	2.8	2.9	2.7	2.8	4.1	3.0	2.7	2.8	2.8	2.9	2.9
Arg	6.1	5.9	5.1	6.3	5.4	5.8	5.9	7.4	7.6	7.7	7.0
AspX	14.5	12.7	14.6	13.4	14.8	14.2	17.2	14.2	14.4	14.2	14.4
GluX	17.2	17.4	18.3	15.7	16.8	16.4	14.7	15.6	15.0	15.3	15.6
Pro	5.2	4.9	5.6	7.1	5.6	6.0	5.0	5.8	5.7	5.8	5.8
Ala	13.6	14.6	13.0	11.9	12.4	12.6	15.0	14.5	14.7	14.3	14.7
Val	10.7	10.7	10.7	10.7	10.4	10.6	9.1	10.0	10.0	10.2	10.0
Leu	12.8	11.6	11.7	11.8	11.5	11.9	11.1	12.1	11.9	12.0	12.4
Tyr	4.8	3.8	3.6	3.7	4.0	4.0	5.0	3.8	4.0	3.9	3.9
Phe	5.6	5.0	4.5	4.9	5.5	5.4	4.8	4.9	4.9	4.8	4.8
Total	99.9	99.9	99.9	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0
					Unstable						
Gly	13.1	13.7	11.7	11.6	11.7	12.9	10.6	12.8	12.4	11.9	12.0
Thr	7.8	7.9	7.8	6.2	7.7	8.2	7.6	7.7	7.6	7.8	7.8
Ser	5.8	6.1	6.3	4.5	7.2	7.1	6.1	6.4	6.3	6.2	6.5
Ileu	10.7	8.4	8.2	8.9	8.1	7.0	6.2	7.5	7.4	7.5	7.5
Met	1.4	4.0	3.4	1.1	3.9	3.7	4.4	4.1	3.9	3.9	4.3
Cys	0.2	0.5	0.7	0.2	0.3	—	0.4	0.5	0.6	0.7	0.9
Total	39.0	40.6	38.1	32.5	38.9	38.9	35.3	39.0	38.2	38.0	39.0
Unknown‡ (%)	0.28	0.62	0.35	0.26	0.38	0.51	0.45	0.47	0.40	0.39	0.31
					Contamination**						
DNA (%)		1.4		1.5		1.4		4.5	1.3	3.9	1.7
RNA (%)		23.3		21.0		12.4		18.6	7.1	11.7	2.4

* GC contents of DNA adopted in this table were taken from reference 1.
† EBM: Enriched broth medium (ref. 7). MM: Minimal medium (ref. 8).
‡ Expressed as a percentage in area of the ninhydrin positive, unknown material over that of the total area.

supernatant (a crude extract) was saved. "Preparation 1" was obtained by heating the crude extract at 100°C for 10 min and then by dialyzing against three liters of distilled water for 24 hr at 4°C with constant stirring. The external water was changed once. A further purification of the crude extract for "preparation 2" was achieved by adding urea to a final concentration of 4.5 *M*, and by stirring the solution at room temperature for 30 min. Ammonium sulfate (0.6 gm per ml) was added to the mixture, and the solution was centrifuged at 12,000 *g* for 20 min at 5°C. The precipitate was washed with 20 ml of 90% ethanol twice, suspended in 10 ml of distilled water, and dialyzed against 3 liters of distilled water for 24 hr at 4°C with constant stirring. The external water was changed once. The resulting preparation is called "preparation 2."

The amounts of protein of preparations 1 and 2 were determined by the Biuret method. The contents of DNA and RNA were determined by indole[9] and orcinol[10] methods, respectively. These are expressed as weight per cent of protein (Tables 1 and 2). On the average, 2.8 per cent of DNA and 16 per cent of RNA were found in the preparations listed in Table 1. For amino acid analyses, the procedure of reference 11 was followed. An aliquot of preparation 2 containing 5 mg of protein was taken in a test tube, and water was added to make the volume 2 ml. Two milliliters of concentrated HCl (12 *N*) was added to it, and the tube was sealed under vacuum after the sample was frozen. The sealed tube was kept in an air-drying oven at 110 ± 1°C for 22 hr. The hydrolyzate was dried and redissolved in 5 ml of a 0.2 *N* sodium citrate buffer, pH 2.2.[11] Minute amounts of carbon appeared after hydrolysis. This was removed by filtration. The amino acid composition was analyzed by a SPINCO Automatic Amino Acid Analyzer.

In all but three of the samples analyzed, unknown ninhydrin positive peaks were quite rare, and they occupied less than 1 per cent of the total peak area. The exceptions were *Serratia mar-*

POSITION OF TOTAL PROTEINS

Aerobacter aerogenes 57	*Serratia marcescens* ATCC 264 58		*Sarcina lutea* ATCC 381 64		*Pseudomonas aeruginosa* ATCC 10197 65		*Alcaligenes faecalis* ATCC 8750 67	*Micrococcus lysodeikticus* ATCC 4698 72		*Tetrahymena pyriformis* 1A 25 1% broth medium
EBM Prep 2	EBM Prep 2	MM Prep 2	EBM Prep 2	EBM Prep 2	EBM Prep 2	MM Prep 2	EBM Prep 2	EBM Prep 2	EBM Prep 2	Prep 2
					Stable					
7.7	7.6	8.2	6.6	7.7	7.2	8.2	9.5	6.5	5.5	11.9
2.5	2.6	2.5	2.8	3.3	3.0	2.7	3.3	2.4	2.1	2.8
6.8	6.3	7.1	9.3	10.4	8.4	8.4	10.2	8.2	6.5	6.1
14.8	16.7	14.5	12.4	11.5	13.2	12.7	12.3	12.1	12.4	16.5
15.6	13.0	15.5	15.8	14.9	15.8	15.6	12.9	16.5	17.6	17.6
5.9	5.0	5.7	6.4	6.3	6.3	6.4	6.5	6.4	6.2	5.1
15.7	15.4	15.5	17.9	17.6	15.5	16.2	14.9	19.8	21.6	10.3
10.2	9.8	10.1	11.0	10.6	10.0	10.0	10.4	11.1	11.8	8.0
12.5	12.5	12.2	10.9	11.4	12.4	12.5	11.8	11.7	11.2	11.7
3.7	5.1	4.0	3.0	2.5	3.5	2.9	3.4	1.7	2.1	4.3
4.5	6.0	4.7	3.9	3.7	4.9	4.4	4.6	3.4	3.0	6.0
99.9	100.0	100.0	100.0	99.9	100.0	99.9	99.9	100.0	100.0	100.0
					Unstable					
12.9	11.9	13.1	14.3	13.9	12.7	12.9	12.7	14.9	15.2	11.0
7.6	6.9	7.4	8.4	6.9	9.2	6.8	6.8	7.7	8.1	7.7
6.7	7.4	6.9	4.5	4.7	6.4	6.3	6.3	4.9	5.8	8.1
7.1	6.1	6.6	4.5	5.6	6.5	6.5	7.0	5.6	5.0	8.8
2.4	4.2	6.0	1.0	1.1	3.2	3.3	4.0	0.1	2.1	1.6
0.5	0.6	0.7	—	—	0.8	0.4	—	—	—	—
37.2	37.1	40.7	32.7	32.2	38.8	36.2	36.8	33.2	36.2	37.2
0.69	1.13	1.11	0.78	0.77	0.48	0.39	0.34	0.87	1.02	0.50
					Contamination**					
4.7		2.6			3.2	2.9				6.4
31.5		18.8			28.2	19.3				21.1

§ ATCC: American Type Culture Collection. KM: A nonsporulating strain. NCTC: National Collection of Type Cultures, London, England. *T. pyriformis* 1A-10602b was given by Dr. D. L. Nanney.
** Expressed as a weight per cent of protein. All available samples were examined and presented here.

cescens grown on either enriched broth medium (1.1%) or minimal medium (1.1%) and one culture of *Micrococcus lysodeikticus* grown on enriched broth medium (1.0%).

Results.—The relative molar content of each amino acid was expressed in the following manner: molar amounts of amino acids, lysine, histidine, arginine, aspartic acid and asparagine, glutamic acid and glutamine, proline, alanine, valine, leucine, tyrosine, and phenylalanine, which are known to be stable and to be well recoverable in the analysis,[12] were summed. Other amino acids including glycine, threonine, serine, isoleucine, methionine, and cysteine are classified as "unstable amino acids." Glycine was not included in the "stable" class, because some glycine is produced by the decomposition of contaminating nucleic acids.[13] It is evident from the data of Tables 1 and 2 that no correlations exist between the extent of contamination with nucleic acid and the observed amino acid composition of the different protein preparations. Subsequently, the amount of each amino acid, both "stable" and "unstable," was expressed by its proportion to the sum of the stable amino acids.[14] The values expressed in this way are presented in Table 1 and Figure 1. The base compositions of DNA of bacteria used are taken from reference 1 and expressed as molar guanine-cytosine content in percentage. The above way of calculation has several advantages which may be listed as follows. (*a*) It obviates the error in estimating the relative amounts of the stable amino acids stem-

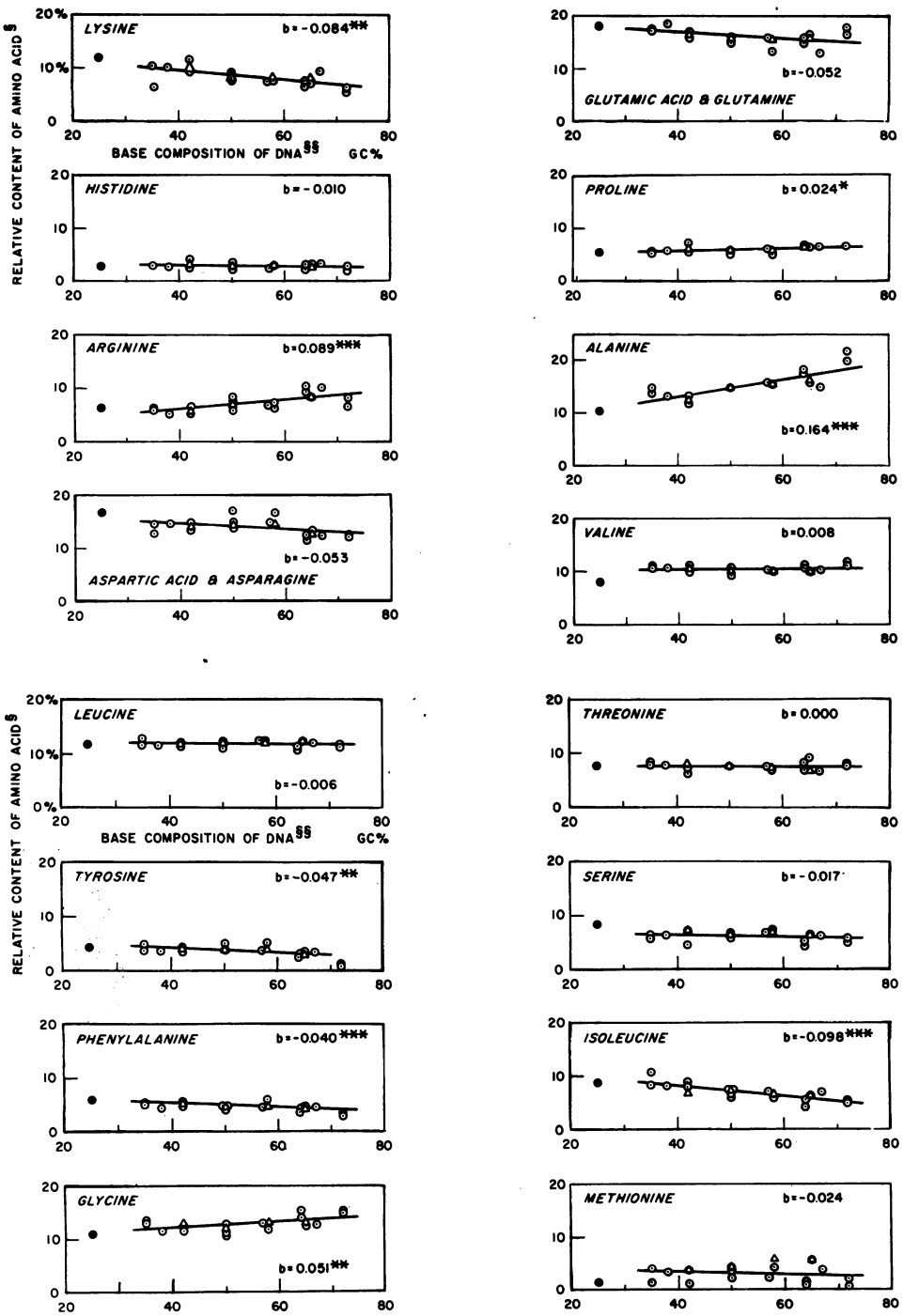

FIG. 1.—Relation between DNA base composition and amino acid composition of protein in bacteria. The relative molar content of each amino acid is expressed as a percentage of the amino acid content over the sum of stable amino acids: lysine, histidine, arginine, aspartic acid plus asparagine, glutamic acid plus glutamine, proline, alanine, valine, leucine, tyrosine, and phenylalanine. The base composition of DNA is expressed as a molar content (%) of guanine plus cytosine (GC). The GC contents of bacterial species adopted in this work are the data of reference 1. Regression coefficient (b). *Tetrahymena* data are not included in the calculation. Levels of significance of b's (or slope) are shown by asterisks.

* Significant ($0.02 < p < 0.05$).
** Highly significant ($0.001 < p < 0.02$).
*** Extremely significant ($p < 0.001$).
⊙ Enriched broth medium.
△ Minimal medium.
● *Tetrahymena pyriformis* (1A-10602 b) in 1 per cent broth medium.

ming from poor recoveries of unstable amino acids. (*b*) Relative contents of all amino acids are not changed after the calculation. (*c*) The sum of the "stable" amino acids is used as the denominator, and these have both positive and negative regression coefficients. Moreover, "stable" amino acids, the sum of which is used as

TABLE 2

AMINO ACID COMPOSITION OF TOTAL PROTEIN OF *E. coli* IN VARIOUS CONDITIONS

Medium Temperature (°C) Growth phase Preparation	EBM 37 Stationary Prep 1	EBM 37 Stationary Prep 2	EBM 37 Stationary Prep 2	EBM 37 Logarithmic Prep 2	EBM 30 Stationary Prep 2	MM 37 Stationary Prep 2	Data of Roberts *et al.* (1957)
			Stable				
Lys	8.9	8.9	8.9	9.3	8.8	8.5	10.0
His	2.8	2.8	2.9	2.6	2.8	2.9	1.3
Arg	7.4	7.6	7.7	8.2	7.3	7.0	7.7
AspX	14.2	14.4	14.2	13.5	14.3	14.4	14.2
GluX	15.6	15.0	15.3	15.8	15.2	15.6	15.1
Pro	5.8	5.7	5.8	5.4	5.7	5.8	6.6
Ala	14.5	14.7	14.3	14.9	14.8	14.7	18.2
Val	10.0	10.0	10.2	11.0	10.3	10.0	7.9
Leu	12.1	11.9	12.0	11.2	11.9	12.4	11.3
Tyr	3.8	4.0	3.9	3.4	4.1	3.9	3.0
Phe	4.9	4.9	4.8	4.6	4.8	4.8	4.7
Total	100.0	100.0	100.0	99.9	99.9	100.0	100.0
			Unstable				
Gly	12.8	12.4	11.9	15.9	12.6	12.0	11.5
Thr	7.7	7.6	7.8	7.6	7.5	7.8	6.7
Ser	6.4	6.3	6.2	7.5	6.3	6.5	8.7
Ileu	7.5	7.4	7.5	7.6	7.5	7.5	6.6
Met	4.1	3.9	3.9	3.2	3.8	4.3	4.8
Cys	0.5	0.6	0.7	0.4	0.6	0.9	2.5
Total	39.0	38.2	38.0	42.2	38.3	39.0	40.8
Unknown* (%)	0.47	0.40	0.39	0.44	0.36	0.31	
			Contamination†				
DNA (%)	4.5	1.3	3.9	0.8	2.1	1.7	
RNA (%)	18.6	7.1	11.7	45.5	9.9	2.4	

* Expressed as a percentage in area of the ninhydrin-positive unknown material over that of the total area.
† Expressed as a weight percentage of protein.

a common denominator, include a majority (13) of the twenty amino acids. As a consequence, the calculation will hardly affect the regression coefficients for the unstable amino acids.

Although there is no special reason to choose linear relations between the GC content of DNA and the amino acid content of protein, a straight line was simulated by least square methods. A summary of the statistical analysis is given in Table 3. Several amino acids show significant regression coefficients, both positive and negative.

The relative amino acid content is quite reproducible, which is indicated by the results on duplicate cultures of most species. The effect of environmental differences on the relative amino acid content of the total protein was examined by changing the composition of medium and the culture temperature and by comparing logarithmic and stationary phases of growth (Tables 1 and 2). No significant differences were observed. The present data on the amino acid composition of *E. coli* protein agree fairly well with those obtained by Roberts *et al.*[15] (Table 2). Selective enrichment of particular amino acids during the protein isolation procedure seems to be unlikely from the data on preparations 1 and 2 in *E. coli*. The relative amounts of amino acids in both fractions are very similar.

TABLE 3

ANALYSIS OF TABLE 1

Amino acid	Average content ($\bar{y}$)	Regression coefficient (b)	Confidence limits of b (5% level)	Probability	Extrapolation of GC content to 0%	Extrapolation of GC content to 100%
			Stable			
Lys	8.46	−0.084	±0.046	0.01–0.001	12.96	4.56
His	2.48	−0.010	±0.015	0.10–0.20	3.38	2.38
Arg	7.23	+0.089	±0.043	<0.001	2.46	11.36
AspX	13.85	−0.053	±0.054	0.10–0.05	16.69	11.39
GluX	15.77	−0.052	±0.050	0.05–0.10	18.14	12.98
Pro	5.89	+0.024	±0.020	0.02	4.60	7.00
Ala	15.30	+0.164	±0.055	<0.001	6.51	22.91
Val	10.39	+0.008	±0.023	0.30–0.50	9.96	10.76
Leu	11.90	−0.006	±0.020	0.50–0.70	12.22	11.62
Tyr	3.63	−0.047	±0.035	0.01–0.02	6.15	1.45
Phe	4.68	−0.040	±0.022	<0.001	6.82	2.82
			Unstable			
Gly	12.80	+0.051	±0.039	0.01–0.02	10.07	15.17
Thr	7.61	+0.0001	±0.027	>0.90	7.60	7.62
Ser	6.11	−0.017	±0.033	0.30–0.50	7.12	5.32
Met	3.07	−0.024	±0.062	0.30–0.50	4.36	1.96
Ileu	7.06	−0.098	±0.032	<0.001	12.31	2.51

Least-square method was used to calculate a and b for a linear relation $y = a + b(x - \bar{x})$. Here y is a relative molar content of an amino acid, and $\bar{y}$, its average, while x and $\bar{x}$ are GC content of DNA and its average. It is noted that a is equal to $\bar{y}$, the average of y. The way of regression analysis used here is described in reference 21. In calculating regression coefficents all but Tetrahymena data were included.

Discussion.—Several interesting features emerge from this investigation. In the first place, no unknown amino acids are found which are correlated to the base compositions of DNA. Secondly, the amino acid composition of the total protein is remarkably invariant to environmental changes as shown in Table 2 for *E. coli* data. For example, minimal versus enriched medium, stages of growth, and culture temperature do not affect the results appreciably. The amino acid composition of the total protein of any species is obviously an average composition of many kinds of protein. The constancy of the over-all composition suggests that no proteins possessing an amino acid composition widely disparate from the average are produced in large proportions under the conditions examined. In this respect it will be remembered that the compositional heterogeneity of DNA of a bacterial species is rather small compared with that of higher organisms.[6] It is possible that the extent of compositional heterogeneity in DNA may reflect that of amino acid composition in protein. Thirdly, individual amino acids appear in distinguishably different molar proportions. For example, histidine, tyrosine, phenylalanine, and methionine are found at distinctly lower levels when compared with other amino acids. This information seems to be extremely important when we realize that the relative content of each amino acid could reflect the frequency of a certain sequence of nucleotides in DNA. Finally, and most significant of all, the present results indicate that there exist several significant correlations between DNA base composition and amino acid composition of protein. According to the nature of correlations, we can classify amino acids into three groups. Among 18 amino acids tested, alanine, arginine, glycine, and proline are positively correlated with the guanine-cytosine (GC) content of DNA. Isoleucine, lysine, aspartic acid plus asparagine, glutamic acid plus glutamine, tyrosine, and phenylalanine are negatively correlated. Histidine, valine, leucine, threonine, serine, and possibly methionine are extremely uniform with no detectable evidence of correlation.

The correlation disclosed here may not, however, represent the true picture for coded protein, since some noncoded proteins may exist. Consequently, the correlations reported here may represent underestimations of the true values. When more data on amino acid composition of different cellular components become available, we will be in a better position to evaluate this evidence in relation to the coding problem. For this reason, any detailed deductions of a coding model from the present data are not attempted in this paper. Nevertheless, some qualitative discussions will be attempted with the hope that they may stimulate further investigations along this line.

(1) The present data seem to support universality of the code among bacteria. The presence of different codes among the bacteria would clearly preclude finding any correlation. Relevant to this issue are the preliminary data on amino acid content of total protein of *Tetrahymena pyriformis* included in Table 1. *T. pyriformis* strain 1A has 25 per cent GC content with a small heterogeneity.[16] The amino acid content of its protein fits remarkably well with the extrapolated lines of the data obtained with the bacteria (see Table 1). This may suggest that the underlying coding is also common to protozoa. (2) The argument that the wide variation of DNA base composition is due to the presence of large amounts of "nonsense" DNA[4, 17] seems to be rather unlikely because of the presence of correlations. This is particularly true of the correlations observed with arginine, tyrosine, and isoleucine, where the extrapolation to 0 or 100 per cent GC content makes the amino acid content almost zero. (3) The "6 keto-6 amino two-symbol code"[4, 18] is not supported by the present data, for in such a system we would expect no correlations. (4) If the coding ratio is universal and relatively small (e. g., 3, 4, 5, 6, etc.), it seems logical to have even numbers, namely 4 or 6, rather than odd numbers to explain the existance of amino acids which show neither positive nor negative correlations. To illustrate this point, let us take triplet and quadruplet nonoverlapping four-symbol codes.[19, 20] Since our analysis is based on the GC content of DNA, two base pairs, A-T and T-A, are indistinguishable, and they are expressed as α-pairs. Similarly G-C and C-G pairs are denoted as γ-pairs. Following this terminology, we can classify the possible triplets into two classes, *exclusive* and *asymmetric*. The triplets of the exclusive class are those which contain only α or γ, and we may designate them by the symbols, α_3 and γ_3. Analogously, asymmetric triplets are ones which have both α and γ but in different numbers, namely $\alpha_2\gamma$ and $\alpha\gamma_2$.

Following a similar argument, we can classify the quadruplets into three classes, exclusive (α_4 and γ_4), asymmetric ($\alpha_3\gamma$ and $\alpha\gamma_3$), and *symmetric* ($\alpha_2\gamma_2$). The symmetric class consists of quadruplets with equal numbers of α's and γ's. Then the frequency of each type of triplet or quadruplet will have correlations of different signs and magnitudes with the GC content of DNA depending on the content of γ in the particular type. Accordingly, the same kinds of correlations will be expected between amino acids in protein corresponding to each triplet (or quadruplet) and the GC content of DNA (Table 4). It is noted that the symmetric class will have no correlation when a linear regression line is estimated. The expected correlations within our GC range of analysis (25 to 75 per cent) are shown in Table 4. If we force our data to this scheme, the best fit is given by assignments similar to that in Table 4.

TABLE 4

NON-OVERLAPPING TRIPLET AND QUADRUPLET CODES AND EXPECTED CORRELATIONS BETWEEN DNA BASE COMPOSITION (GC CONTENT) AND AMINO ACID CONTENTS OF PROTEINS

Triplets

Classes	Type	Expected regression to GC content of DNA
Exclusive	α_3	− −
	γ_3	+ +
Asymmetric	$\alpha_2\gamma$	−
	$\alpha\gamma_2$	+

Quadruplets

Classes	Types	Expected regression to GC content of DNA	Amino acids classified
Exclusive	α_4	− −	Lysine, isoleucine
	γ_4	+ +	Arginine, alanine
Symmetric	$\alpha_2\gamma_2$	0	Histidine, valine, leucine, threonine, serine, methionine
Asymmetric	$\alpha_3\gamma$	−	Aspartic acid + asperagine, glutamic acid + glutamine, tyrosine, phenylalanine
	$\alpha\gamma_3$	+	Glycine, proline

α: Either A-T or T-A. γ: Either G-C or C-G. A: adenine. T: thymine. G: guanine. C: cytosine. +, + +: positively correlated. −, − −: negatively correlated. 0: no correlation.

Needless to say, the validity of these issues must be subjected to further experimental tests. Nevertheless, the results are encouraging for our faith in the existence of a coding mechanism amenable to an ultimate resolution.

Summary.—Amino acid composition of the bulk protein isolated from 11 bacterial species whose DNA base compositions vary from 35 to 72 per cent in guanine-cytosine content has been examined. The following points are of interest.

1. The amino acid composition of the total protein is remarkably invariant to environmental changes.

2. Individual amino acids appear in distinguishably different molar proportions with rather uniform values.

3. There exist several significant correlations between DNA base composition and amino acid composition of protein. Among 18 amino acids tested, alanine, arginine, glycine, and proline are positively correlated with guanine-cytosine content of DNA. Isoleucine, lysine, aspartic acid plus asparagine, glutamic acid plus glutamine, tyrosine, and phenylalanine are negatively correlated. Histidine, valine, leucine, threonine, serine, and possibly methionine are extremely uniform with no detectable evidence of correlation.

The results obtained were discussed in relation to the coding problem.

I wish to thank Mrs. C. W. Luker and Mr. R. Igarashi for their help with the amino acid analysis and Mrs. M. E. McCarty for her aid in preparation of the manuscript. I am grateful to Drs. S. Spiegelman and A. D. Hershey for reading and criticizing the manuscript. This work was supported by a grant from the National Science Foundation (G-15080).

[1] Lee, K. Y., R. Wahl, and E. Barbu, *Ann. Inst. Pasteur*, **91**, 212 (1956).
[2] Belozersky, A. N., and A. S. Spirin, *Nature*, **182**, 111 (1958).
[3] Rolfe, R., and M. Meselson, these PROCEEDINGS, **45**, 1039 (1959).
[4] Sueoka, N., J. Marmur, and P. Doty, *Nature*, **183**, 1429 (1959).
[5] Sueoka, N., these PROCEEDINGS, **45**, 1480 (1959).
[6] Sueoka, N., *J. Mol. Biol.*, **3**, 31 (1961).

[7] Tissières, A., J. D. Watson, D. Schlessinger, and B. R. Hollingworth, *J. Mol. Biol.*, **1**, 221 (1959).
[8] Sistrom, W. R. See Folsome, C. E., *Genetics*, **45**, 1111 (1960).
[9] Keck, K., *Arch. Biochem. Biophys.*, **63**, 446 (1956).
[10] Dische, Z., and K. Schwarz, *Mikrochim. Acta*, **2**, 13 (1937).
[11] Spackmann, D. H., W. H. Stein, and S. Moore, *Anal. Chem.*, **30**, 1190 (1958).
[12] Tsugita, A., and H. Fraenkel-Conrat, these PROCEEDINGS, **46**, 636 (1960).
[13] Fraser, D., *J. Biol. Chem.*, **227**, 711 (1957).
[14] In order to express the molar content of each amino acid as the percentage of the total amino acids, the value listed in Tables 1 and 2 should be divided by a factor of 1.4.
[15] Roberts, R. B., D. B. Cowie, P. H. Abelson, E. T. Bolton, and R. J. Britten, *Studies of Biosynthesis in Escherichia coli*, Publ. 607 of the Carnegie Institute of Washington (1957), p. 28.
[16] Sueoka, N., and D. L. Nanney (in preparation).
[17] Crick, F. H. C., *Brookhaven Symp. Biol.*, No. 12, 35 (1959).
[18] Sinsheimer, R. L., *J. Mol. Biol.*, **1**, 218 (1959).
[19] Crick, F. H. C., J. S. Griffith, and L. E. Orgel, these PROCEEDINGS, **43**, 416 (1957).
[20] Golomb, S. W., Welch, L. R., and Delbrück, M., *Biol. Medd. Dan. Vid. Selsk.*, **23**, 9 (1958).
[21] Mather, K., *Statistical Analysis in Biology* (London: Methuen & Co., Ltd., 1951), pp. 109–128.

ON THE SPECIES SPECIFICITY OF ACCEPTOR RNA AND ATTACHMENT ENZYMES

BY SEYMOUR BENZER AND BERNARD WEISBLUM

DEPARTMENT OF BIOLOGICAL SCIENCES, PURDUE UNIVERSITY

Communicated June 26, 1961

One of the steps in protein biosynthesis appears to be the attachment of each amino acid to a specific acceptor (SRNA) molecule. According to the adaptor hypothesis, each SRNA molecule would then fit to a specific complementary base sequence on a linear RNA template, specifying the sequence of amino acids in the resultant protein.[1, 2] An adaptor molecule thus could have two specificities: one recognizing the correct amino acid and activating enzyme; the other, the proper position on the template. The correctness of the amino-acid sequence therefore would depend upon the precision and constancy of the adaptors. However, the structures of the enzymes and adaptors are presumably under the genetic control of the organism and might be subject to heritable modifications. It is therefore conceivable that one or both ends of an adaptor might change sufficiently to cause occasional errors and, in the long run, an alteration of the genetic code might evolve. This notion, prompted by genetic observations[3] which suggested that mutation of a bacterium might modify its translation of genetic information, lead to the present comparison of the specificities of the acceptor RNA and activating enzymes of different organisms.

Several differences in specificity have been reported previously. Berg *et al.*[4] demonstrated that SRNA from *Escherichia coli* contains two distinguishable acceptors for methionine. An enzyme prepared from yeast could attach methionine to one of these, while the enzyme from *E. coli* could attach to both. Webster found, in pig liver, a difference between the nuclear and cytoplasmic attachment enzymes for alanine. Rendi and Ochoa[6] noted that, for leucine, the enzymes in